

사례보고

“코로나-19: 우리의 기억”: 코로나바이러스 감염증과 사회변화에 대한 디지털 아카이브

“COVID-19: Our Memory”: A Digital Archive for Social Changes
caused by SARS-CoV-2

김학래(Haklae Kim)

E-mail: haklaekim@cau.ac.kr

중앙대학교 문헌정보학과 교수



논문접수 2020-10-26
최초심사 2020-10-30
게재확정 2020-11-04

ORCID

Haklae Kim
<https://orcid.org/0000-0002-2616-421X>

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

초 록

코로나바이러스감염증은 인류사회가 경험하지 못한 커다란 충격과 생활양식의 급속한 변화를 만들고 있다. 비대면 사회는 감염병 확산을 예방하기 위한 과정에서 보편화된 사례이다. 코로나바이러스감염증으로 인한 사회적 영향은 광범위하다. 정부의 정책, 개인 정보보호, 정보기술 등 다양한 이슈가 사회 전반에 영향을 주고 있다. 동시에 관련 사건과 이슈가 신속하고 빠르게 변하기 때문에 사실 정보를 추적하고 기록하는 것이 어렵다. 코로나-19와 실시간성 정보를 효과적으로 기술하기 위한 방안은 무엇일까?

“코로나-19: 우리의 기억” 프로젝트는 코로나바이러스감염증에 대한 사회문화적 영향을 가치중립적으로 기록하기 위한 시도이다. 주요 사건과 이슈를 분야별로 수집하고, 중립적인 관점으로 핵심이벤트를 기록하며, 모든 기록을 탐색할 수 있도록 디지털 아카이브로 구축한다. 프로젝트를 통해 수집, 구축한 모든 데이터, 소스코드, 시각화를 포함하는 애플리케이션은 모두 공개하여 새로운 협업을 이끌어내고 있다.

ABSTRACT

In light of SARS-CoV-2's significant impact, human society has experienced rapid changes in lifestyle that it has not yet experienced before. One way this virus has influenced people's lives is the emergence of the zero-contact society, an initiative for preventing the spread of infectious diseases. As can be seen, the social impact of COVID-19 is widespread. Various issues, such as those about government policy, personal information protection, and health care, are affecting society as a whole. At the same time, factual information is difficult to track and record because of the rapid and transient nature of related events and issues. As such, a method of effectively describing COVID-19 and real-time information is necessary.

The “COVID-19: Our Memory” project is an attempt to record the sociocultural impact of the coronavirus infection. This project collects major events and issues classified into several subjects, records those events from a neutral point of view, and develops a digital archive so that all records are accessible. All the data collected and built through the project, the application, including the source code and visualization, are all published to bring about new opportunities for collaboration.

Keywords: 디지털 아카이브, COVID-19, 데이터 분석, 데이터 시각화
Digital archives, COVID-19, Data analysis, Data visualization

1. 서론

코로나바이러스감염증(코로나19)은 코로나 바이러스로 인해 발생하는 감염질환이다. 중국에서 최초 확진환자가 보고된 후(2019년 12월), 감염병은 아시아, 유럽, 미국 등 여러 대륙에서 동시다발적으로 빠르게 확산되었다(Bedford et al., 2020, p. 1014). 세계보건기구(WHO: World Health Organization)는 세계적 대유행(pandemic)을 선언하고(WHO, 2020년 3월), 범세계적 차원에서 감염병의 확산 방지와 예방을 강조한다(Jebril, 2020, p. 2784). 우리나라는 2월 이후 확진환자가 급격히 증가했으나, 정부의 기민한 대응, 광범위한 사회적 거리두기, 수준 높은 시민의식을 통해 바이러스 확산을 통제하고 있다고 평가받고 있다(Park, 2020, p. 2129). 코로나바이러스감염증은 보건의료에 한정하지 않고 정치, 경제, 문화, 과학기술 등 인류의 모든 생활에 급격한 변화를 만들고 있다. 비대면 사회(zero-contacted society)는 코로나바이러스감염증에 대응하기 위한 과정에서 보편화된 대표적 사례이다.

기록의 관점에서 보면, 코로나바이러스감염증은 사회 전반에 영향을 미치는 사건과 이슈를 객관적으로 기록함에 있어 새로운 질문을 던지고 있고, 동시에 기록의 중요성을 제고시키는 계기로 작용하고 있다. 역사적 사건의 기록을 해석하는 것과 다르게 현재 진행형의 사건은 여러 가지 고려사항이 있다(Dixon, 2020). 첫째, 가치 중립적이고 객관적으로 기록해야 한다. 예컨대, 마스크(mask)는 코로나19 예방을 위해 효과적이라는 주장과 보건의료적 효과가 미흡하다는 상반된 의견이 있고, 국가와 문화적 상황에 따라 마스크 착용에 대한 권고가 다르다(Feng, 2020, p. 435). 세계보건기구는 팬데믹을 선언하기 이전까지 마스크 착용에 대해 권고하지 않았으나, 범세계적 확산과 함께 수정된 권고안을 발표한다. 국내의 상황을 고려하면, 마스크와 관련된 논쟁은 조금 더 다양하다. 국내에서 마스크는 코로나19 예방을 위해 필수품으로 인식된다. 그러나, 급속한 감염 확산으로 마스크가 부족함에 따라, 정부는 ‘마스크 5부제’라는 제한적 공급 정책을 시행하게 된다(Moon, 2020, p. 652). 마스크 부족의 원인은 다양하다. 시민의 불안감이 증폭되어 사재기와 같은 가수요를 만든 측면도 있고, 정부의 정책 혼선이 수요와 공급의 불균형을 만들었다는 비판도 존재한다. 기록의 차원에서 이런 상황은 함께 중립적으로 기록되어야 한다. 둘째, 신속하고, 윤리적이어야 한다. 코로나바이러스감염증은 국가를 넘어 전 세계적 재난으로 다양한 사건과 이슈가 빠르게 생산되고 상황에 따라 급격한 변화가 이루어진다(An & Tang, 2020, p. 792). 이런 과정에서 감염병에 대한 데이터, 정부의 발표와 대응자료는 주기에 관계없이 수정 또는 삭제되어 완결된 자료 중심으로 보존되는 경향이 있다. 더불어, 확진자 동선 등 개인정보에 대한 법적 해석의 범위를 넘나드는 사례에 대해 균형 있게 기록해야 한다. 예를 들어, 이태원 집단감염으로 인한 성소수자의 정보, 제주도를 여행한 모녀의 개인정보, 8·15 시위 참석자의 검사결과 공개는 공익 목적의 개인정보 이용과 인원 보호 사이의 민감한 사례이다. 확진환자의 동선정보는 수집할 수 있으나 해당 법률에 따라 삭제될 수 있도록 윤리적 기준이 필요하다. 마지막으로, 디지털 환경에서 탐색과 활용이 가능해야 한다. 모든 기록은 웹 환경에서 탐색하고 활용할 수 있어야 한다. 코로나바이러스감염증에 대한 정보는 디지털 환경에서 생산되고 소비된다. 확진환자에 대한 통계, 긴급재난지원금, 마스크 5부제 등 주요 이슈와 대응 과정은 전자화된 정보를 포함하고 있다. 따라서, 모든 기록은 디지털 환경을 기반으로 자유롭게 탐색하고, 재사용할 수 있는 방식을 지원해야 한다.

2. 추진 과정

‘코로나19: 우리의 기억’ 프로젝트¹⁾는 코로나바이러스감염증으로 발생한 국가 전반에 걸친 사건과 영향을 기록하고, 기록된 자원을 재사용하여 새로운 가치를 발현시키는 목표를 갖고 있다. 팀. 케일리(Team.Cayley)는 프로젝

1) <http://okfn.kr/projects/covid-19-our-memory/>

트 수행을 위해 조직되었고, 자발적 참여 의사가 있는 중앙대학교 재학생으로 구성되어 있다. 1기 활동은 2020년 5월에 시작되어 8월에 완료되었고, 2020년 10월부터 2기 활동이 진행되고 있다.

코로나-19: 우리의 기억
Home 목표 주제 포트폴리오 구성원 소통 언어 ▼

어떤 주제를 포함하나

코로나-19의 감염 및 확산 현황을 분석하고, 사회 전반에 커다란 영향을 미치거나 일상 생활의 변화를 이끌었던 이슈를 기록한다.

| | | |
|--|--|---|
| <p>01</p> <p>정부 정책 및 대응</p> <p>코로나-19 예방 및 감염 확산 방지를 위한 정부의 움직임</p> | <p>02</p> <p>경제</p> <p>코로나-19가 미친 경제적 영향 및 주요 분야</p> | <p>03</p> <p>교육</p> <p>코로나-19 확산에 따른 국내 특수, 유치원, 초중고 교 등교수업 조정 현황</p> |
| <p>04</p> <p>문화</p> <p>오프라인 문화산업의 위기(영화, 공연, 관광), 언택트 산업의 기회(도서-출판, OTT)</p> | <p>05</p> <p>사회</p> <p>청와대 청원으로 보는 코로나-19 이슈와 사회 불평등</p> | <p>06</p> <p>환경</p> <p>코로나-19가 환경에 미친 영향과 포스트 코로나의 그린 뉴딜</p> |

무엇을 제공하나

코로나-19와 관련된 국내의 데이터, 분석 및 시각화를 위한 소스 코드, 정부의 보도자료 등 수집 및 구축된 모든 자료를 공개하고 CC-BY 2.0 KR 라이선스로 공개한다.

데이터 수집

주요 주제별로 수집한 데이터

정제 및 구축 데이터

감염 현황, 주요 분야별 수집, 정제 및 구축된 모든 데이터

분석 코드

주제별 분석한 소스 코드 (파이썬)

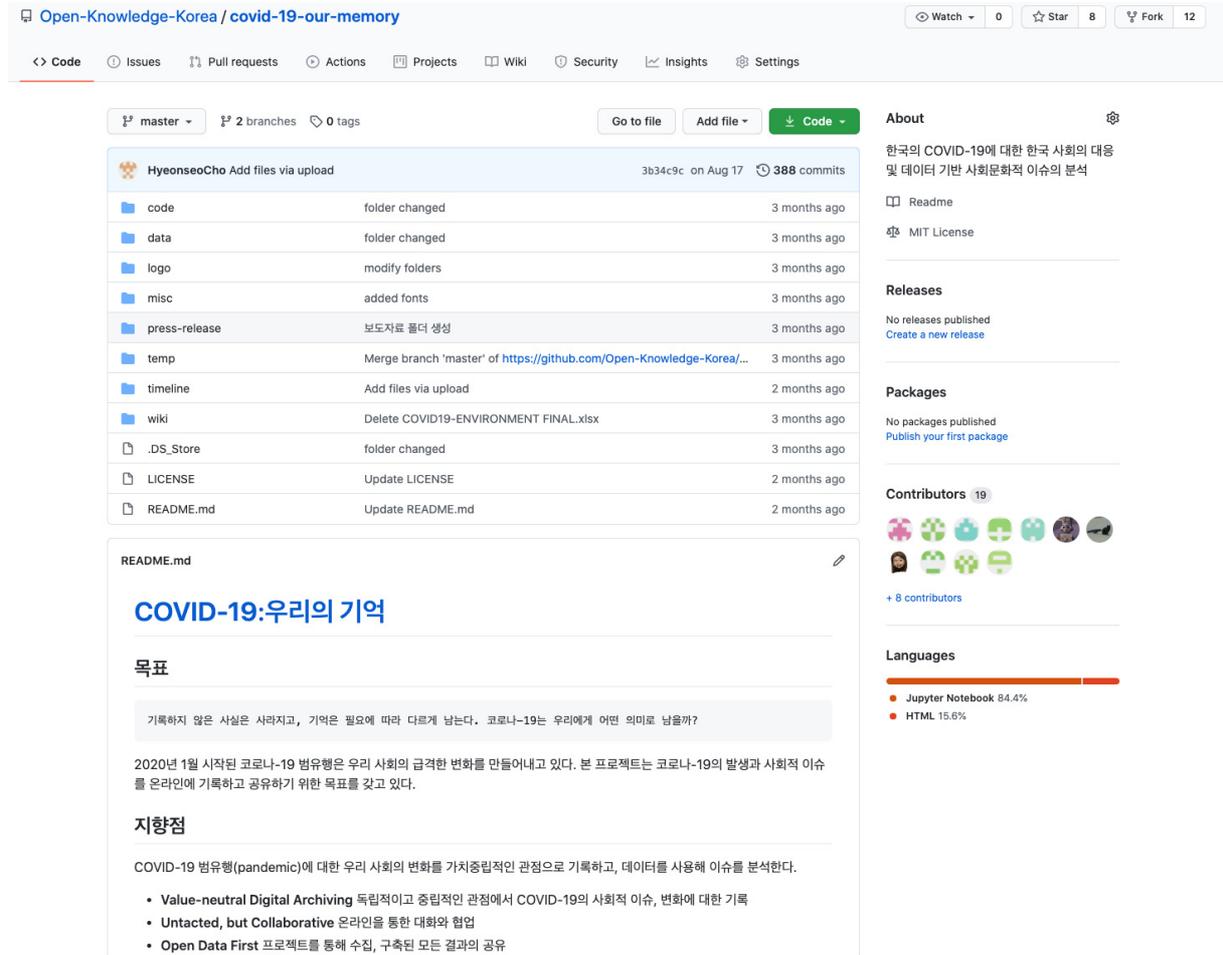
데이터 시각화

코로나-19 관련 데이터의 시각화

〈그림 1〉 “코로나-19: 우리의 기억” 프로젝트의 웹사이트: 프로젝트의 목표와 포트폴리오는 포함되어 있고, 모든 자료는 웹 자원으로 형식으로 탐색할 수 있다.

‘코로나19: 우리의 기억’ 프로젝트는 디지털 아카이브에 사실 정보를 기록하고, 기록의 검증과 재현을 위해 다음의 세 가지 원칙을 정의하고 있다.

- (1) 가치중립적 디지털 기록(Value-neutral Archiving)
- (2) 지속 가능한 협업 모델(Zero-contacted, but Collaborative)
- (3) 오픈 데이터 원칙(Open Data First)



<그림 2> 프로젝트 깃허브 저장소: 수집한 데이터, 분석을 위한 소스코드, 분석 결과를 공유

첫째, 모든 기록은 구성원의 논의를 통해 중립적인 관점으로 기술한다. 디지털 아카이브의 주요 주제는 정부 정책과 대응, 경제, 교육, 문화, 사회, 환경으로 구분한다. 의료, 정치, 노동, 인권, 여성 등 세분화할 수 있는 주제가 있지만, 개별 이슈의 지속성과 빈도를 고려해 세분화하지 않고 주요 주제에 포함시킨다. 세부 주제별 팀은 매주 이벤트(사건 또는 이슈)를 수집한다. <표 1>과 같이 2월부터 7월까지 수집한 이벤트는 약 835개이다. 정부정책과 대응 관련된 이벤트가 상대적으로 빈도가 높고, ‘문화’는 영화, 관광, 도서출판 등 다양한 주제를 포함한 특성이 반영되어 있다. 핵심 이벤트는 수집된 이벤트에서 사회적 영향과 파급력을 고려하여 선정한다. 선정 과정은 주제 팀의 1차 논의와 전체 구성원이 참여한 2차 검증을 통해 진행한다. 주제별로 균형 있는 분포를 고려하여, 핵심 이벤트는 주제별 필수 이벤트를 5~6개로 제한하고, 논의를 통해 이벤트를 추가하는 방식으로 진행한다. 현재 선정된 핵심 이벤트는 약 200개이다.

둘째, ‘코로나19: 우리의 기억’ 프로젝트는 사회적 거리두기가 엄격하게 지속되는 기간에 진행되었다. 프로젝트의 모든 과업이 비대면 환경에서 이루어지기 때문에 새로운 방식의 소통과 협력이 필요하다. 실제 개별 팀과 전체 구성원의 정기회의는 온라인 환경에서 비대면으로 진행되었다. 슬랙(Slack)은 구성원 사이의 정보 공유를 위한 목적으로 활용하고, 구글 드라이브(Google Drive)는 프로젝트에 활용되는 모든 자원을 공유하는데 사용되었다. 매주 진행되는 정기회의는 줌(Zoom)을 사용했다. 특히, 정기회의에 참여하지 못한 구성원을 위해, 모든 회의록과 회의 영상은 구성원에게 공유되었다.

마지막으로 ‘코로나19: 우리의 기억’은 오픈데이터를 원칙으로 진행된다. 단편적이고 일시적인 프로젝트를 넘어, 프로젝트의 목표는 지속가능한 협력을 이끌어내는 것을 지향한다. 프로젝트를 통해 수집된 원시 데이터, 기록된 모든 데이터, 데이터 분석을 위해 개발된 모든 소스코드는 모든 대중에게 공개한다. 모든 자료는 깃허브(Github)를 통해 공유하고, 특수한 사례를 제외하고 크리에이티브 커먼즈-저작권 표시 대한민국 라이선스를 적용한다.

<표 1> 주제별 수집 이벤트 현황

| 구분 | 정부정책 | 경제 | 교육 | 문화 | 사회 | 환경 |
|------|------|----|----|-----|----|----|
| 수집규모 | 402 | 87 | 79 | 140 | 73 | 54 |

3. 디지털 아카이브와 서비스

‘코로나19: 우리의 기억’ 프로젝트의 주요 활동은 감염병에 대한 사회과학적 문제 정의와 이를 해결하기 위한 공학적 접근, 그리고 구성원의 집단지성을 통한 문제의 해결 과정을 포함하고 있다. 디지털 아카이브는 코로나19와 관련된 다양한 통계 데이터, 뉴스 데이터, 정부의 보도자료와 이를 분석한 소스코드, 시각화 서비스를 융합적 산출물을 포함한다. 모든 자료는 깃허브의 ‘covid-19-our-memory’ 저장소²⁾에서 확인할 수 있다.

3.1 데이터셋

코로나19와 관련된 국내외 데이터 1,067이 포함되어 있다. 확진환자, 마스크 수출입 동향 등 정부가 발표한 공식적인 데이터는 우선적으로 활용한다. 다만, 정확성이 결여되거나 미흡한 데이터는 새롭게 구축해 공개한다. 예컨대, 마스크 공급에 대한 데이터는 식품의약품안전처에서 제공하고 있지만, 한글문서와 이미지 형식으로 제공되기 때문에 발표된 자료를 기계가 처리할 수 있는 형식으로 구축한다.

3.2 데이터 분석

주요 주제에서 정의한 문제는 데이터 분석으로 종합적인 상황을 판단하는 시도를 하였다. 주제별로 정의한 문제는 크게 6가지를 포함한다.

- 마스크 부족의 원인(<https://tinyurl.com/yx8wlvmm>)
- 등교수업 조정의 원칙과 실제(<https://tinyurl.com/y57xyvry>)
- 문화, 관광 산업의 영향(<https://tinyurl.com/yyzpg6sa>)
- 재난 불평등: 재난은 평등한가?(<https://tinyurl.com/yy8xyj74>)
- 산업별 고용 동향(<https://tinyurl.com/y3fjuxpc>)
- 대기오염, 일회용품, 의료폐기물이 환경에 주는 영향(<https://tinyurl.com/y2y9ec6s>)

2) <https://github.com/Open-Knowledge-Korea/covid-19-our-memory>

방역수칙 단계별 전환 참고지표

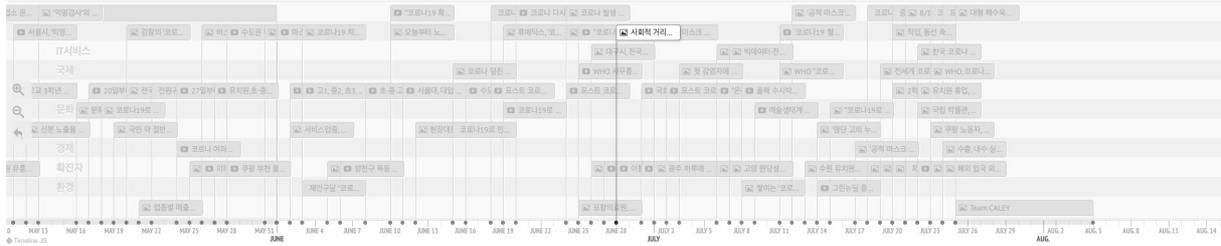
| 위험도 평가 참고지표 | 사회적 거리두기 | | |
|----------------------------|---|--|--|
| | 1단계 (생활속거리두기) | 2단계 | 3단계 |
| 구별(최근 3주간) | 50명 미만 | 50~100명 미만 | 100~200명 이상, 1주 2회 이상 발생 |
| 일일 확진환자 수(평균) (지역사회 환자 증가) | 50명 미만 | 50~100명 미만 | 100~200명 이상, 1주 2회 이상 발생 |
| 감염경로 불분명 사례 비율 | 5% 미만 | --- | 급격한 증가 확인 |
| 관리중인 집단발생 현황(건) | 감소 또는 0% 이상 | 지속적 증가 | 급격한 증가 확인 |
| 방역팀 내 관리 비율(%) | 증가 또는 80% 이상 | --- | --- |
| 유형의 심각성 및 기준 | 지역발생 및 집단 발생 수를 이해하기 위하여 유형이 확인된 사례를 분석 | 통상적인 집단발생이 발생 가능할 수준을 초과해 주체, 시·군·구에서 코로나19 유행이 발생하여 관리가 필요함 | 지역사회에서 다수의 집단 발생이 발생하여 코로나19 감염의 급속도로 확산하는 것으로 추정됨 |

*1주일 2회 이상 일일확진자 100명 이상 증가

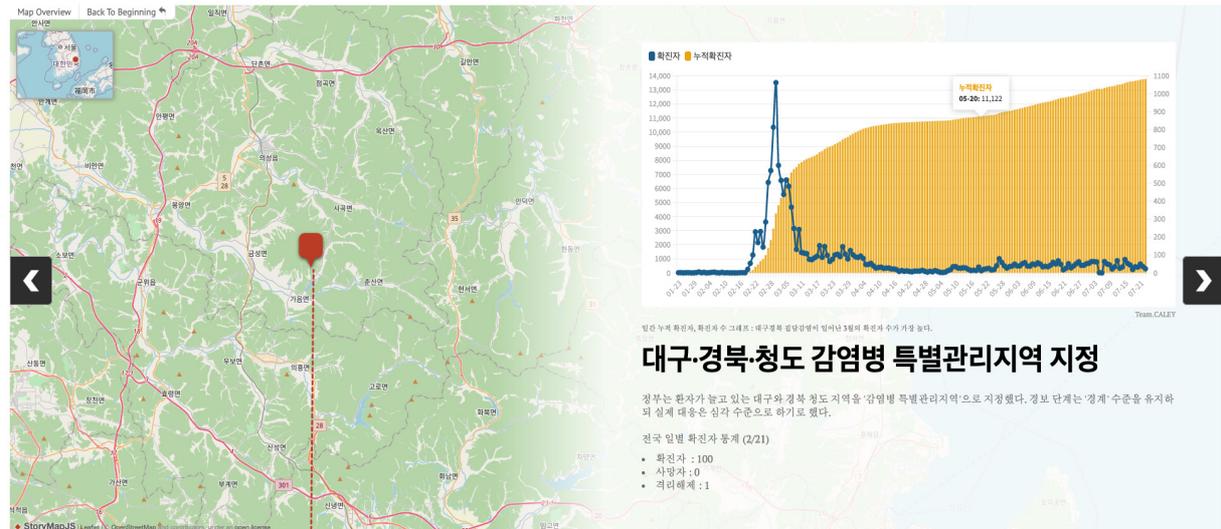
사회적 거리두기 3단계로 재 정비

정부가 사회적 거리두기 강도에 따라 1-3단계로 구분하고 각 단계에 적용할 지침을 재 정비했다. 신종 코로나바이러스 감염증(코로나19) 하루 신규 확진자 수가 50명 미만이면 1단계로, 이 숫자가 50-100명으로 늘면 2단계를 적용한다. 일일 확진자 수가 100명 이상으로 치솟아 대유행으로 번지면 3단계를 적용해 모든 모임과 다중이용시설, 학교 등 영을 중단한다.

출처: news.kmib.co.kr/article/view.a...



〈그림 3〉 코로나19 타임라인: 주제별로 선정한 핵심 이벤트를 시간에 따라 보여주는 기능



〈그림 4〉 코로나19 스토리맵: 집단감염이 발생된 주요 지역을 시각적으로 보여주는 기능

분석 결과 페이지는 문제에 대한 배경, 데이터 수집방법, 분석 방법에 대해 자세하게 기술하고 있다. 설정한 문제에 따라 데이터 분석과 시각화 방법은 다양하다. 예를 들어, 마스크 품귀 현상은 정부의 마스크 공급 정책에 대한 데이터 수집과 시각화를 제공한다. 반면, 문화, 재난 불평등, 고용에 대한 주제는 파이썬으로 데이터 분석과 시각화를 수행했다. 데이터 분석에 활용된 소스코드는 파이썬 노트북 형식으로 제공해 분석 결과를 쉽게 확인할 수 있다.

3.3 데이터 시각화

핵심 이벤트로 선정된 항목은 뉴스 데이터, 멀티미디어 자료, 분석 데이터, 정부의 보도자료를 활용해 하나의

이벤트로 재구성된다. 즉, 코로나19 디지털 아카이브의 단일 이벤트는 텍스트, 멀티미디어, 동적 시각화 정보를 함께 연계한 복합 정보이다. 복합 정보를 구성하기 위해, 코로나19 관련 통계 데이터는 인터랙티브 차트(Interactive chart)로 시각화한다. 이와 같은 형식은 데이터를 동적으로 제공하고, 동시에 웹 페이지에 이식할 수 있기 때문에 정보의 재사용을 높일 수 있다. 뉴스와 같은 텍스트 정보는 약 150 단어로 요약하고, 관련 이벤트의 설명으로 사용되며 원본 출처의 주소를 함께 제공한다. 복합적인 정보를 제공하기 위한 시각화는 크게 타임라인과 스토리맵으로 구성한다. 먼저, 코로나19 타임라인은 국내에서 첫 번째 확진 환자의 발생부터 주요 사건을 시간적으로 구성해 탐색할 수 있는 기능을 제공한다. 스토리맵은 국내에서 발생한 집단감염을 공간적 관점으로 재구성하고, 시간의 순서와 지역의 감염확산을 시각적으로 보여준다.

4. 향후 계획

‘코로나-19: 우리의 기억’ 프로젝트는 코로나바이러스감염증에 대한 사회문화적 영향을 가치중립적으로 기록하기 위한 시도이다. 2020년 1월부터 시작된 감염병의 확산은 우리 사회 전반에 걸쳐 커다란 충격을 주었고 동시에 변화의 필요성을 제기하고 있다. 기록 관점에서 보면, 국가 재난으로 인한 사회적 이슈와 급격하고 광범위한 영향을 미치는 사건을 효과적으로 기록하기 위한 방안에 대해 논의가 필요하다. 또한 디지털화된 사건의 가치중립성을 제고하기 위해 데이터의 정확성이 매우 중요하다. 코로나19 관련 데이터는 정부가 관리하고 있지만, 범정부 차원의 데이터 거버넌스의 부재로 일관된 형식으로 기록되지 못하고 있다. 외국인 유입 환자, 확진환자의 동선정보, 지방자치단체의 데이터 공개 방식 등 데이터를 기록하기 위한 거버넌스 체계 수립이 필요하다.

‘코로나-19: 우리의 기억’의 2기 활동은 일반 대중의 참여와 협력을 이끌어내는 것을 목표로 한다. 수집, 정제된 데이터는 캐글(Kaggle)에 공유하고, 정부의 데이터를 수집해 기계가 처리할 수 있는 형식으로 변환하기 위해 데이터 파이프라인 구축을 계획하고 있다. 또한 디지털 아카이브에 포함된 모든 데이터는 RiC-O, Schema.org 어휘를 적용하여 서로 다른 유형의 데이터를 연계하고, 보편적인 접근을 위한 질의 인터페이스를 구축할 예정이다.

참고문헌

- An BY & Tang S-Y. (2020). Lessons From COVID-19 Responses in East Asia: Institutional Infrastructure and Enduring Policy Instruments. *The American Review of Public Administration*, 50(6-7), 790-800.
<https://doi.org/10.1177%2F0275074020943707>
- Bedford, Juliet, Enria, Delia, Giesecke, Johan, Heymann, David, Ihekweazu, Chikwe, Kobinger, Gary, Lane, H., Memish, Ziad, Oh, Myoung-don, Sall, Amadou, Schuchat, Anne, Ungchusak, Kumnuan, & Wieler, Lothar (2020). COVID-19: Towards Controlling of a Pandemic. *The Lancet*, 395, 1015-1017. [https://doi.org/10.1016/S0140-6736\(20\)30673-5](https://doi.org/10.1016/S0140-6736(20)30673-5).
- Dixon, Jennifer (2020). Documenting the Pandemic: Libraries Launch COVID-19 Archival Projects.
<https://www.libraryjournal.com/?detailStory=Documenting-Pandemic-Libraries-Launch-COVID-19-Archival-Projects-archives>
- Feng, Shuo, Shen, Chen, Xia, Nan, Song, Wei, Fan, Mengzhen, & Cowling, Benjamin (2020). Rational Use of Face Masks in the COVID-19 Pandemic. *The Lancet Respiratory Medicine*, 8, 434-436.
[https://doi.org/10.1016/S2213-2600\(20\)30134-X](https://doi.org/10.1016/S2213-2600(20)30134-X).
- Jebriil, Nadia (2020). World Health Organization Declared a Pandemic Public Health Menace: A Systematic Review of the

Coronavirus Disease 2019 “COVID-19.” *International Journal of Psychosocial Rehabilitation*, 24, 2784-2795.
<https://dx.doi.org/10.2139/ssrn.3566298>.

Moon, M. (2020). Fighting Against COVID-19 with Agility, Transparency, and Participation: Wicked Policy Problems and New Governance Challenges. *Public Administration Review*, 651-656. <https://doi.org/10.1111/puar.13214>

Park S., Choi G.J., & Ko H. (2020). Information Technology-Based Tracing Strategy in Response to COVID-19 in South Korea—Privacy Controversies. *JAMA*, 323(21), 2129-2130. <https://doi.org/10.1001/jama.2020.6602>